PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
B09B 3/00, C12N 1/00, 5/10, 9/02, 15/53, 15/63, C12P 1/00, 7/02

A1

(11) International Publication Number: **WO 99/08812**

(43) International Publication Date: 25 February 1999 (25.02.99)

(21) International Application Number: PCT/US98/16979

(22) International Filing Date: 17 August 1998 (17.08.98)

(30) Priority Data:
60/056,754 20 August 1997 (20.08.97) US

(71) Applicant: THE UNIVERSITY OF ROCHESTER [US/US]; Office of Technology Transfer, 518 Hylan Building, Rochester, NY 14627 (US).

(72) Inventors: JONES, Jeffrey, P.; 400 N.W. Maryland Court, Pullman, WA 99163 (US). SHIMOJI, Miyuki; Apartment 4C, 229 Conant Road, Rochester, NY 14623 (US).

(74) Agents: WEYAND, Karla, M. et al.; Nixon, Hargrave, Devans & Doyle LLP, Clinton Square, P.O. Box 1051, P.O. Box 1051, Rochester, NY 14603 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: FUNCTIONAL BACTERIAL/MAMMALIAN CYTOCHROME P450 CHIMERA

(57) Abstract

The present invention is directed to a chimeric DNA molecule which includes a first DNA molecule encoding a portion of a full length bacterial P450 protein and a second DNA molecule fused to the first DNA molecule and encoding a portion of a full length mammalian P450 protein. The chimeric DNA molecule encodes a fusion protein which is active and soluble in aqueous liquid. A further aspect of the present invention is directed to the fusion protein encoded by the chimeric DNA molecule. The fusion protein is useful in bioremediation processes and also can be used to hydroxylate a compound to be oxidized.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgía	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

FUNCTIONAL BACTERIAL/MAMMALIAN CYTOCHROME P450 CHIMERA

The subject matter of this application was made with support from the United States Government National Institutes of Health Grant No. GM624(PPG), ES060062, and ES05407. The Government may have certain rights.

This application claims benefit of U.S. Provisional Patent Application Serial No. 60/056,754, filed August 20, 1997, which is hereby incorporated by reference.

FIELD OF THE INVENTION

The present invention relates to a functional bacterial/mammalian cytochrome P450 chimera.

BACKGROUND OF THE INVENTION

Cytochrome P450 ("P450") is a term used for a widely distributed group of unique heme proteins which form carbon monoxide complexes with a major absorption band at wavelengths around 450 nm. These proteins are enzymes which carry out oxidations involved in biosynthesis and catabolism of specific cell or body components, and in the metabolism of foreign substances entering organisms. Oxygenating enzymes such as P450 appear to be fundamental cellular constituents in most forms of aerobic organisms. The activation of molecular oxygen and incorporation of one of its atoms into organic compounds by these enzymes are reactions of vital importance not only for biosynthesis, but also for metabolic activation or inactivation of foreign agents such as drugs, food preservatives and additives, insecticides, carcinogens and environmental pollutants.

In eukaryotic systems P450, and P450 dependent enzymes are known to act on such xenobiotics and pharmaceuticals as phenobarbitol, antipyrine, haloperidol and prednisone. Known substrates of environmental importance include compounds such as DDT, and a variety of polychlorinated biphenyls and polyaromatic hydrocarbons, as well as other halogenated compounds, including halobenzenes and chloroform.

Hexamethylphosphoramide ("HMPA") is a compound that was used heavily by industry in the mid-1970's in the production of aramid fibers and as a general solvent. HMPA is a known carcinogen and has been found to be one of the contaminants at various industrial and chemical waste sites. Studies focusing on the mammalian biodegradation of HMPA are few but it has been found that microsomal P450 isolated from rat liver and nasal mucosa will demethylate HMPA. (Longo et al., Toxicol. Lett. 44:289 (1988)).

In microbial systems, cytochrome P450 is known to oxidize many of the same xenobiotic substrates as in eukaryotic systems and thus can be targeted as possible indicators for the presence of toxic compounds in the environment. One of the earliest reports of xenobiotic transformation was by the bacterium *Streptomyces giseus* which is known to contain the gene for the expression of cytochrome P450. This transformation involved the convention of mannosidostreptomycin to streptomycin. (Sariaslani et al., Developments in Industrial Microbiology 30:161 (1989)). Since then, these reactions have been observed with compounds ranging from simple molecules such as benzene to complex alkaloids (such as vindoline and dihydrovindolin, codein, steroids, and xenobiotics such as phenylhydrazine, ajmaline and colchine. (Sariaslani et al., Developments in Industrial Microbiology 30:161 (1989)).

Genetically engineered microorganisms with the ability to express the P450 gene offer several potential advantages. Such microorganisms might be designed to express precisely engineered enzymatic pathways that can more efficiently or rapidly degrade specific chemicals. Development efforts are aimed largely at chemicals that are toxic or recalcitrant to naturally occurring bacterial degradation.

It has also been shown that enzyme-substrate interactions can be a dominant feature of P450 mediated reactions. (Paulsen et al., Methods in Enzymology, 272:337-46 (1996)). To date no three-dimensional structure of a mammalian P450 enzyme is available despite the use of special expression vectors (Sandhu et al., "Expression of Modified Cytochrome P450 2C10 (2C9) in *Escherichia coli*, Purification, and Reconstitution of Catalytic Activity," Arch. Biochem. Biophys., 306:443-450 (1993); Haining et al., "Allelic Variants of Human Cytochrome

P4502C9: Baculovirus-mediated Expression, Purification, Structural Characterization, Substrate Stereoselectivity, and Prochiral Selectivity of the Wild-Type and I359L Mutant Forms," Arch. Biochem. Biophys., 333:447-458 (1996); Waterman, M.S., "Heterologus Expression of Mammalian P450 Enzymes," Advances Enzymol., 68:37-66 (1994)) and peptitergents to improve solubility. (Sueyoshi et al., "Molecular Engineering of Microsomal P4502a-4 to a Stable, Water-Soluble Enzyme," Arch. Biochem. Biophys., 322:265-271 (1995)). In contrast, the crystal structures of a number of cytosolic bacterial P450s have been determined. These include $P450_{cam}$, $P450_{bm3}$, $P450_{terp}$, and $P450_{eryF}$. (Poulos et al., "The 2.6-Δ Crystal Structure of *Psudomonas putida* Cytochrome P-450," J. Biol. Chem., 260:16122-16130 (1985); Poulos et al., "High-Resolution Crystal Structure P450cam," J. Mol. Biol., 195:685-700 (1987); Ravichandran et al., "Crystal Structure of Hemeprotein Domain of P450BM-3, a Prototype for Microsomal P450's," Science, 261:731-736 (1993); Hasemann et al., "Crystal Structure and Refinement of Cytochrome $P450_{terp}$ at 2.3 Δ Resolution," J. Mol. Biol., 1169-1185 (1994); Haseman et al., "Structure and Function of Cytochrome P450: A Comparative Analysis of Three Crystal Structures," Structure, 3:41-62 (1995); Cupp-Vickery et al., "Preliminary Crystallographic Analysis of an Enzyme Involved in Erythromycin Biosynthesis: Cytochrome $P450_{eryF}$," Proteins, 20:197-201 (1994)). Since no detailed structural information has been obtained for a mammalian P450 enzyme, all attempts to determine the effect of enzyme-substrate interactions have used the crystal structures from the soluble bacterial P450 enzymes. (Cupp-Vickery et al., "Preliminary Crystallographic Analysis of an Enzyme Involved in Erythromycin Biosynthesis: Cytochrome $P450_{eryF}$," Proteins, 20:197-201 (1994); Paulsen et al., Methods in Enzymology, 272:337-46 (1996)). While homology models can be constructed for the membrane-bound mammalian enzymes based on the bacterial enzymes, the very low sequence identities (<20%) mean that any resulting model is of low resolution. In fact, no information directly shows that mammalian and bacterial enzymes are structurally related.

The present invention is directed to overcoming the deficiencies of the prior art by forming a P450 protein which is soluble and active in aqueous liquid.

SUMMARY OF THE INVENTION

The present invention is directed to a chimeric DNA molecule which includes a first DNA molecule encoding a portion of a full length bacterial P450 protein and a second DNA molecule fused to the first DNA molecule and encoding a portion of a full length mammalian P450 protein. The chimeric DNA molecule encodes a fusion protein which is active and soluble in aqueous liquid.

Another aspect of the present invention relates to a fusion protein which includes a portion of a bacterial P450 protein and a portion of a mammalian P450 protein fused to the portion of a bacterial P450 protein. The fusion protein is active and soluble in aqueous liquid.

In addition, the chimeric DNA molecule of the present invention is useful in the bioremediation of an environmental pollutant. The method involves contacting the environmental pollutant with the fusion protein under conditions effective to effect bioremediation.

In addition, the fusion protein is useful in a process of hydroxylating a compound to be oxidized. This involves contacting the compound to be oxidized with the fusion protein under conditions effective to hydroxylate the compound to be oxidized.

This fusion protein has a number of advantages over the native enzymes. For example, since the protein is soluble, it will lend itself to structural elucidation by X-ray crystallography. This is very important in terms of protein design. In addition, a protein is provided, as well as the potential to design a number of proteins, that can be readily expressed in a soil bacteria that will use the bacterial reductases. This has implications for both bioremediation and the biosynthesis of organic compounds. The fusion protein is an important step forward in allowing the use of the less restrictive mammalian active site architecture, which should allow for the design of more diversely functional proteins. Further, since the chimera uses bacterial enzyme that are present in soil bacteria, it can be expressed in this bacterial vector and the bacteria applied to the soil. This obviates the need for coexpression of mammalian reductases while still retaining the prefered active site geometry of the mammalian enzymes.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A is a model of the chimeric structure of the present invention. The blue region is from $P450_{cam}$ and the red region is from CYP2C9. The chimera contains 3 substrate recognition sites from $P450_{cam}$ and 3 from CYP2C9. Figure 1B shows the construction of a fused plasmid of $P450_{cam}$ and CYP2C9.

Figure 2A is a CO-reduced differential spectrum of the fusion protein of the present invention. The preparation used corresponds to lane 2 in Figure 2B. Figure 2B shows an SDS-polyacrylamide gel electrophoresis of the chimera of the present invention expressed in *E. coli*. Lanes 1 and 2 show the fusion protein and lane 3 and 4 show $P450_{cam}$ wild-type. Lane 1, 105,000g supernatant (3μg protein); lane 2, eluate from a hydroxylaapatite column (1.5 μg protein); lane 3, 105,000g supernatant (3 μg protein); lane 4, eluate from hydroxylapatite column (2.2 μg protein); lane 5, molecular marker. The gel was stained with Coomassie Brilliant Blue R250.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to a chimeric DNA molecule which includes a first DNA molecule encoding a portion of a full length bacterial P450 protein and a second DNA molecule fused to the first DNA molecule and encoding a portion of a full length mammalian P450 protein. The chimeric DNA molecule encodes a fusion protein which is active and soluble aqueous liquid. This chimeric DNA molecule can have the nucleotide sequence corresponding to SEQ. ID. No. 1 as follows:

```
atgacgactg aaaccataca aagcaacgcc aatcttgccc ctctgccacc ccatgtgcca 60
gagcacctgg tattcgactt cgacatgtac aatccgtcga atctgtctgc cggcgtgcag 120
gaggcctggg cagttctgca agaatcaaac gtaccggatc tggtgtggac tcgctgcaac 180
ggcggacact ggatcgccac tcgcggccaa ctgatccgtg aggcctatga agattaccgc 240
cacttttcca gcgagtgccc gttcatccct cgtgaagccg gcgaagccta cgacttcatt 300
cccacctcga tggatccgcc cgagcagcgc cagtttcgtg cgctggccaa ccaagtggtt 360
ggcatgccgg tggtggataa gctggagaac cggatccagg agctggcctg ctcgctgatc 420
gagagcctgc gcccgcaagg acagtgcaac ttcaccgagg actacgccga acccttcccg 480
```

atacgcatct tcatgctgct cgcaggtcta ccggaagaag atatcccgca cttgaaatac 540

ctaacggatc agatgacccg tccggatggc agcatgacct tcgcagaggc caaggaggcg 600

ctctacgact atctgatacc gatcatcgag caacgcaggc agaagccggg aatgaacaac 660

cctcaggact ttattgattg cttcctgatg aaaatggaga aggaaaagca caaccaacca 720

tctgaattta ctattgaaag cttggaaaac actgcagttg acttgtttgg agctgggaca 780

gagacgacaa gcacaaccct gagatatgct ctccttctcc tgctgaagca cccagaggtc 840

acagctaaag tccaggaaga gattgaacgt gtgattggca gaaaccggag cccctgcatg 900

caagacagga gccacatgcc ctacacagat gctgtggtgc acgaggtcca gagatacatt 960

gaccttctcc ccaccagcct gccccatgca gtgacctgtg acattaaatt cagaaactat 1020

ctcattccca agggcacaac catattaatt tccctgactt ctgtgctaca tgacaacaaa 1080

gaatttccca acccagagat gtttgaccct catcactttc tggatgaagg tggcaatttt 1140

aagaaaagta aatacttcat gcctttctca gcaggaaaac ggatttgtgt gggagaagcc 1200

ctggccggca tggagctgtt tttattcctg acctccattt tacagaactt taacctgaaa 1260

tctctggttg acccaaagaa ccttgacacc actccagttg tcaatggatt tgcctctgtg 1320

ccgcccttct accagctgtg cttcattcct gtctga 1356

The chimeric DNA molecule, corresponding to SEQ. ID. No. 1, encodes a fusion protein which includes a portion of a full length bacterial P450 protein and a portion of a full length mammalian P450 protein fused to the portion of the full length bacterial P450 protein. The fusion protein is active, soluble, and can have the amino acid sequence of SEQ. ID. No. 2 as follows:

Asn Leu Ala Pro Leu Pro Pro His Val Pro Glu His Leu Val Phe Asp
1 5 10 15

Phe Asp Met Tyr Asn Pro Ser Asn Leu Ser Ala Gly Val Gln Glu Ala
20 25 30

Trp Ala Val Leu Gln Glu Ser Asn Val Pro Asp Leu Val Trp Thr Arg
35 40 45

Cys Asn Gly Gly His Trp Ile Ala Thr Arg Gly Gln Leu Ile Arg Glu
50 55 60

Ala Tyr Glu Asp Tyr Arg His Phe Ser Ser Glu Cys Pro Phe Ile Pro
65 70 75 80

Arg Glu Ala Gly Glu Ala Tyr Asp Phe Ile Pro Thr Ser Met Asp Pro
85 90 95

Pro Glu Gln Arg Gln Phe Arg Ala Leu Ala Asn Gln Val Val Gly Met
100 105 110

Pro Val Val Asp Lys Leu Glu Asn Arg Ile Gln Glu Leu Ala Cys Ser
115 120 125

Leu Ile Glu Ser Leu Arg Pro Gln Gly Gln Cys Asn Phe Thr Glu Asp
130 135 140

Tyr Ala Glu Pro Phe Pro Ile Arg Ile Phe Met Leu Leu Ala Gly Leu
145 150 155 160

Pro Glu Glu Asp Ile Pro His Leu Lys Tyr Leu Thr Asp Gln Met Thr
165 170 175

Arg Pro Asp Gly Ser Met Thr Phe Ala Glu Ala Lys Glu Ala Leu Tyr
180 185 190

Asp Tyr Leu Ile Pro Ile Ile Glu Gln Arg Arg Gln Lys Pro Gly Asn
195 200 205

Asn Pro Gln Asp Phe Ile Asp Cys Phe Leu Met Lys Met Glu Lys Glu
210 215 220

Lys His Asn Gln Pro Ser Glu Phe Thr Ile Glu Ser Leu Glu Asn Thr
225 230 235 240

Ala Val Asp Leu Phe Gly Ala Gly Thr Glu Thr Thr Ser Thr Thr Leu
245 250 255

Arg Tyr Ala Leu Leu Leu Leu Leu Lys His Pro Glu Val Thr Ala Lys
260 265 270

Val Gln Glu Glu Ile Glu Arg Val Ile Gly Arg Asn Arg Ser Pro Cys
275 280 285

Met Gln Asp Arg Ser His Met Pro Tyr Thr Asp Ala Val Val His Glu
290 295 300

Val Gln Arg Tyr Ile Asp Leu Leu Pro Thr Ser Leu Pro His Ala Val
305 310 315 320

Thr Cys Asp Ile Lys Phe Arg Asn Tyr Leu Ile Pro Lys Gly Thr Thr
325 330 335

Ile Leu Ile Ser Leu Thr Ser Val Leu His Asp Asn Lys Glu Phe Pro
340 345 350

Asn Pro Glu Met Phe Asp Pro His His Phe Leu Asp Glu Gly Gly Asn
355 360 365

Phe Lys Lys Ser Lys Tyr Phe Met Pro Phe Ser Ala Gly Lys Arg Ile
370 375 380

Cys Val Gly Glu Ala Leu Ala Gly Met Glu Leu Phe Leu Phe Leu Thr
385 390 395 400

Ser Ile Leu Gln Asn Phe Asn Leu Lys Ser Leu Val Asp Pro Lys Asn
405 410 415

```
Leu Asp Thr Thr Pro Val Val Asn Gly Phe Ala Ser Val Pro Pro Phe
            420                 425                 430

Tyr Gln Leu Cys Phe Ile Pro Val His His His His His His
        435                 440                 445
```

The chimeric DNA molecule contains 10 to 90 percent, preferably about 50 percent, of the first DNA molecule and 90 to 10 percent, preferably 50 percent of the second DNA molecule. It is particularly desirable for the first and second DNA molecules to be fused together at a location where the encoded fusion protein lacks secondary structure. This is where there are no interactions due to hydrogen bonds (e.g., at random coils) in the components of the fusion protein.

The chimeric DNA molecule is prepared from a DNA molecule encoding a full length mammalian P450 protein where a portion of that DNA molecule encoding a full length mammalian P450 protein is replaced with a DNA molecule encoding a homologous portion of a full length bacterial P450 protein. This involves replacing all amino acids prior to a random coil between G- and H-helices in the full length mammalian P450 protein with a homologous portion of the full length bacterial P450 protein.

The fusion protein of the present invention is characterized by being soluble. Since eucaryotic P450 proteins are membrane bound, they are insoluble. By contrast, bacterial P450 proteins are soluble. Thus, in the fusion protein of the present invention, the bacterial P450 protein portion imparts its characteristic solubility to the mammalian P450 protein portion.

Another characteristic of the fusion protein of the present invention is that it is active. P450 activity can be defined as the oxidation of a substrate. The most important of these reactions is the removal of a hydrogen atom and replacing it with a hydroxyl group. This reaction is illustrated, for example, by the following:

$$RCH_3 + P450 \rightarrow RCH_2OH$$

where the protein turns a hydrocarbon into an alcohol. Such a reaction is called a hydroxylation reaction. Such reactions are also illustrated in Poulos, "Modeling of

Mammalian P450s on Basis of $P450_{cam}$ X-ray Structure," Methods in Enzymology, 206:11-30 (1991), which is hereby incorporated by reference.

Suitable mammalian P450 proteins include 1A, 2B, 2C, 2D, and 3A families of cytochrome P450 and CYP2C9. CYP2C9, which is particularly preferred, has an amino acid sequence of SEQ. ID. No. 3 as follows:

```
Met Asp Ser Leu Val Val Leu Val Leu Cys Leu Ser Cys Leu Leu Leu
 1               5                  10                  15

Leu Ser Leu Trp Arg Gln Ser Ser Gly Arg Gly Lys Leu Pro Pro Gly
            20                  25                  30

Pro Thr Pro Leu Pro Val Ile Gly Asn Ile Leu Gln Ile Gly Ile Lys
        35                  40                  45

Asp Ile Ser Lys Ser Leu Thr Asn Leu Ser Lys Val Tyr Gly Pro Val
    50                  55                  60

Phe Thr Leu Tyr Phe Gly Leu Lys Pro Ile Val Val Leu His Gly Tyr
65                  70                  75                  80

Glu Ala Val Lys Glu Ala Leu Ile Asp Leu Gly Glu Glu Phe Ser Gly
                85                  90                  95

Arg Gly Ile Phe Pro Leu Ala Glu Arg Ala Asn Arg Gly Phe Gly Ile
            100                 105                 110

Val Phe Ser Asn Gly Lys Lys Trp Lys Glu Ile Arg Arg Phe Ser Leu
        115                 120                 125

Met Thr Leu Arg Asn Phe Gly Met Gly Lys Arg Ser Ile Glu Asp Arg
    130                 135                 140

Val Gln Glu Glu Ala Arg Cys Leu Val Glu Glu Leu Arg Lys Thr Lys
145                 150                 155                 160

Ala Ser Pro Cys Asp Pro Thr Phe Ile Leu Gly Cys Ala Pro Cys Asn
                165                 170                 175

Val Ile Cys Ser Ile Ile Phe His Lys Arg Phe Asp Tyr Lys Asp Gln
            180                 185                 190

Gln Phe Leu Asn Leu Met Glu Lys Leu Asn Glu Asn Ile Lys Ile Leu
        195                 200                 205

Ser Ser Pro Trp Ile Gln Ile Cys Asn Asn Phe Ser Pro Ile Ile Asp
    210                 215                 220

Tyr Phe Pro Gly Thr His Asn Lys Leu Leu Lys Asn Val Ala Phe Met
225                 230                 235                 240

Lys Ser Tyr Ile Leu Glu Lys Val Lys Glu His Gln Glu Ser Met Asp
                245                 250                 255
```

```
Met Asn Asn Pro Gln Asp Phe Ile Asp Cys Phe Leu Met Lys Met Glu
            260                 265                 270

Lys Glu Lys His Asn Gln Pro Ser Glu Phe Thr Ile Glu Ser Leu Glu
        275                 280                 285

Asn Thr Ala Val Asp Leu Phe Gly Ala Gly Thr Glu Thr Thr Ser Thr
    290                 295                 300

Thr Leu Arg Tyr Ala Leu Leu Leu Leu Leu Lys His Pro Glu Val Thr
305                 310                 315                 320

Ala Lys Val Gln Glu Glu Ile Glu Arg Val Ile Gly Arg Asn Arg Ser
                325                 330                 335

Pro Cys Met Gln Asp Arg Ser His Met Pro Tyr Thr Asp Ala Val Val
            340                 345                 350

His Glu Val Gln Arg Tyr Ile Asp Leu Leu Pro Thr Ser Leu Pro His
        355                 360                 365

Ala Val Thr Cys Asp Ile Lys Phe Arg Asn Tyr Leu Ile Pro Lys Gly
    370                 375                 380

Thr Thr Ile Leu Ile Ser Leu Thr Ser Val Leu His Asp Asn Lys Glu
385                 390                 395                 400

Phe Pro Asn Pro Glu Met Phe Asp Pro His His Phe Leu Asp Glu Gly
                405                 410                 415

Gly Asn Phe Lys Lys Ser Lys Tyr Phe Met Pro Phe Ser Ala Gly Lys
            420                 425                 430

Arg Ile Cys Val Gly Glu Ala Leu Ala Gly Met Glu Leu Phe Leu Phe
        435                 440                 445

Leu Thr Ser Ile Leu Gln Asn Phe Asn Leu Lys Ser Leu Val Asp Pro
    450                 455                 460

Lys Asn Leu Asp Thr Thr Pro Val Val Asn Gly Phe Ala Ser Val Pro
465                 470                 475                 480

Pro Phe Tyr Gln Leu Cys Phe Ile Pro Val
                485                 490
```

The DNA molecule encoding CYP2C9 has the nucleotide sequence of SEQ. ID. No. 4 as follows:

```
gaaggcttca atggattctc ttgtggtcct tgtgctctgt ctctcatgtt tgcttctcct 60

ttcactctgg agacagagct ctgggagagg aaaactccct cctggcccca ctcctctccc 120

agtgattgga aatatcctac agataggtat taaggacatc agcaaatcct taaccaatct 180

ctcaaaggtc tatggccctg tgttcactct gtattttggc ctgaaaccca tagtggtgct 240
```

gcatggatat gaagcagtga aggaagccct gattgatctt ggagaggagt ttctggaag 300

aggcattttc ccactggctg aaagagctaa cagaggattt ggaattgttt tcagcaatgg 360

aaagaaatgg aaggagatcc ggcgtttctc cctcatgacg ctgcggaatt ttgggatggg 420

gaagaggagc attgaggacc gtgttcaaga ggaagcccgc tgccttgtgg aggagttgag 480

aaaaaccaag gcctcaccct gtgatcccac tttcatcctg ggctgtgctc cctgcaatgt 540

gatctgctcc attattttcc ataaacgttt tgattataaa gatcagcaat ttcttaactt 600

aatggaaaag ttgaatgaaa acatcaagat tttgagcagc ccctggatcc agatctgcaa 660

taatttttct cctatcattg attacttccc gggaactcac aacaaattac ttaaaaacgt 720

tgcttttatg aaaagttata ttttggaaaa agtaaaagaa caccaagaat caatggacat 780

gaacaaccct caggacttta ttgattgctt cctgatgaaa atggagaagg aaaagcacaa 840

ccaaccatct gaatttacta ttgaaagctt ggaaaacact gcagttgact tgtttggagc 900

tgggacagag acgacaagca caaccctgag atatgctctc cttctcctgc tgaagcaccc 960

agaggtcaca gctaaagtcc aggaagagat tgaacgtgtg attggcagaa accggagccc 1020

ctgcatgcaa gacaggagcc acatgcccta cacagatgct gtggtgcacg aggtccagag 1080

atacattgac cttctcccca ccagcctgcc ccatgcagtg acctgtgaca ttaaattcag 1140

aaactatctc attcccaagg gcacaaccat attaatttcc ctgacttctg tgctacatga 1200

caacaaagaa tttcccaacc cagagatgtt tgaccctcat cactttctgg atgaaggtgg 1260

caatttaag aaaagtaaat acttcatgcc tttctcagca ggaaaacgga tttgtgtggg 1320

agaagccctg gccggcatgg agctgttttt attcctgacc tccattttac agaactttaa 1380

cctgaaatct ctggttgacc caaagaacct tgacaccact ccagttgtca atggatttgc 1440

ctctgtgccg cccttctacc agctgtgctt cattcctgtc tgaagaagag cagatggcct 1500

ggctgctgct gtgcagtccc tgcagctctc tttcctctgg ggcattatcc atctttcact 1560

atctgtaatg ccttttctca cctgtcatct cacattttcc cttccctgaa gatctagtga 1620

acattcgacc tccattacgg agagtttcct atgtttcact gtgcaaatat atctgctatt 1680

ctccatactc tgtaacagtt gcattgactg tcacataatg ctcatactta tctaatgttg 1740

agttattaat atgttattat taaatagaga aatatgattt gtgtattata attcaaaggc 1800

atttcttttc tgcatgttct aaataaaaag cattattatt tgctg 1845

Suitable bacterial P450 proteins include $P450_{cam}$, $P450_{bm3}$, $P450_{terp}$, and $P450_{eryF}$. These proteins are described in Poulos et al., "The 2.6-Δ Crystal Structure of *Psudomonas putida* Cytochrome P-450," J. Biol. Chem., 260:16122-16130 (1985); Poulos et al., "High-Resolution Crystal Structure P450cam," J. Mol. Biol., 195:685-700 (1987); Ravichandran et al., "Crystal Structure of Hemeprotein

Domain of P450BM-3, a Prototype for Microsomal P450's," Science, 261:731-736 (1993); Hasemann et al., "Crystal Structure and Refinement of Cytochrome $P450_{terp}$ at 2.3 Δ Resolution," J. Mol. Biol., 1169-1185 (1994); Haseman et al., "Structure and Function of Cytochrome P450: A Comparative Analysis of Three Crystal Structures," Structure, 3:41-62 (1995); Cupp-Vickery et al., "Preliminary Crystallographic Analysis of an Enzyme Involved in Erythromycin Biosynthesis: Cytochrome $P450_{eryF}$," Proteins, 20:197-201 (1994), which are hereby incorporated by reference. Of these, $P450_{cam}$ is particularly preferred. $P450_{cam}$ has an amino acid sequence of SEQ. ID. No. 5 as follows:

```
Asn Leu Ala Pro Leu Pro Pro His Val Pro Glu His Leu Val Phe Asp
 1               5                  10                  15

Phe Asp Met Tyr Asn Pro Ser Asn Leu Ser Ala Gly Val Gln Glu Ala
            20                  25                  30

Trp Ala Val Leu Gln Glu Ser Asn Val Pro Asp Leu Val Trp Thr Arg
        35                  40                  45

Cys Asn Gly Gly His Trp Ile Ala Thr Arg Gly Gln Leu Ile Arg Glu
    50                  55                  60

Ala Tyr Glu Asp Tyr Arg His Phe Ser Ser Glu Cys Pro Phe Ile Pro
65                  70                  75                  80

Arg Glu Ala Gly Glu Ala Tyr Asp Phe Ile Pro Thr Ser Met Asp Pro
                85                  90                  95

Pro Glu Gln Arg Gln Phe Arg Ala Leu Ala Asn Gln Val Val Gly Met
            100                 105                 110

Pro Val Val Asp Lys Leu Glu Asn Arg Ile Gln Glu Leu Ala Cys Ser
        115                 120                 125

Leu Ile Glu Ser Leu Arg Pro Gln Gly Gln Cys Asn Phe Thr Glu Asp
    130                 135                 140

Tyr Ala Glu Pro Phe Pro Ile Arg Ile Phe Met Leu Leu Ala Gly Leu
145                 150                 155                 160

Pro Glu Glu Asp Ile Pro His Leu Lys Tyr Leu Thr Asp Gln Met Thr
                165                 170                 175

Arg Pro Asp Gly Ser Met Thr Phe Ala Glu Ala Lys Glu Ala Leu Tyr
            180                 185                 190

Asp Tyr Leu Ile Pro Ile Ile Glu Gln Arg Arg Gln Lys Pro Gly Thr
        195                 200                 205
```

```
Asp Ala Ile Ser Ile Val Ala Asn Gly Gln Val Asn Gly Arg Pro Ile
    210                 215                 220

Thr Ser Asp Glu Ala Lys Arg Met Cys Gly Leu Leu Leu Val Gly Gly
225                 230                 235                 240

Leu Asp Thr Val Val Asn Phe Leu Ser Phe Ser Met Glu Phe Leu Ala
                245                 250                 255

Lys Ser Pro Glu His Arg Gln Glu Leu Ile Glu Arg Pro Glu Arg Ile
            260                 265                 270

Pro Ala Ala Cys Glu Glu Leu Leu Arg Arg Phe Ser Leu Val Ala Asp
        275                 280                 285

Gly Arg Ile Leu Thr Ser Asp Tyr Glu Phe His Gly Val Gln Leu Lys
    290                 295                 300

Lys Gly Asp Gln Ile Leu Leu Pro Gln Met Leu Ser Gly Leu Asp Glu
305                 310                 315                 320

Arg Glu Asn Ala Cys Pro Met His Val Asp Phe Ser Arg Gln Lys Val
                325                 330                 335

Ser His Thr Thr Phe Gly His Gly Ser His Leu Cys Leu Gly Gln His
            340                 345                 350

Leu Ala Arg Arg Glu Ile Ile Val Thr Leu Lys Glu Trp Leu Thr Arg
        355                 360                 365

Ile Pro Asp Phe Ser Ile Ala Pro Gly Ala Gln Ile Gln His Lys Ser
    370                 375                 380

Gly Ile Val Ser Gly Val Gln Ala Leu Pro Leu Val Trp Asp Pro Ala
385                 390                 395                 400

Thr Thr Lys Ala Val
                405
```

The DNA molecule encoding $P450_{cam}$ has the nucleotide sequence of SEQ. ID. No. 6 as follows:

```
ctgcaggatc gttatccgct ggccgatctg atcacccagc gtttttccat cgacgaggcc 60

agcaaggcac ttgaactggt caaggcagga gcactgatca aacccgtgat cgactccact 120

ctttagccaa cccgcgttcc aggagaacaa caacaatgac gactgaaacc atacaaagca 180

acgccaatct tgcccctctg ccaccccatg tgccagagca cctggtattc gacttcgaca 240

tgtacaatcc gtcgaatctg tctgccggcg tgcaggaggc ctgggcagtt ctgcaagaat 300

caaacgtacc ggatctggtg tggactcgct gcaacggcgg acactggatc gccactcgcg 360

gccaactgat ccgtgaggcc tatgaagatt accgccactt ttccagcgag tgcccgttca 420

tccctcgtga agccggcgaa gcctacgact tcattcccac ctcgatggat ccgcccgagc 480
```

agcgccagtt tcgtgcgctg gccaaccaag tggttggcat gccggtggtg gataagctgg 540
agaaccggat ccaggagctg gcctgctcgc tgatcgagag cctgcgcccg caaggacagt 600
gcaacttcac cgaggactac gccgaaccct tcccgatacg catcttcatg ctgctcgcag 660
gtctaccgga agaagatatc ccgcacttga aatacctaac ggatcagatg acccgtccgg 720
atggcagcat gaccttcgca gaggccaagg aggcgctcta cgactatctg ataccgatca 780
tcgagcaacg caggcagaag ccgggaaccg acgctatcag catcgttgcc aacggccagg 840
tcaatgggcg accgatcacc agtgacgaag ccaagaggat gtgtggcctg ttactggtcg 900
gcggcctgga tacggtggtc aatttcctca gcttcagcat ggagttcctg gccaaaagcc 960
cggagcatcg ccaggagctg atcgagcgtc ccgagcgtat tccagccgct tgcgaggaac 1020
tactccggcg cttctcgctg gttgccgatg gccgcatcct cacctccgat tacgagtttc 1080
atggcgtgca actgaagaaa ggtgaccaga tcctgctacc gcagatgctg tctggcctgg 1140
atgagcgcga aaacgcctgc ccgatgcacg tcgacttcag tcgccaaaag gtttcacaca 1200
ccacctttgg ccacggcagc catctgtgcc ttggccagca cctggcccgc cgggaaatca 1260
tcgtcaccct caaggaatgg ctgaccagga ttcctgactt ctccattgcc ccgggtgccc 1320
agattcagca caagagcggc atcgtcagcg gcgtgcaggc actccctctg gtctgggatc 1380
cggcgactac caaagcggta taaacacatg ggagtgcgtg ctaagtgaac gcaaacgaca 1440
acgtggtcat cgtcggtacc ggactggctg gcgttgaggt cgccttcggc ctgcgcgcca 1500
gcggctggga aggcaatatc cggttggtgg gggatgcgac ggtaattccc catcacctac 1560
caccgctatc caaagctt 1578

The protein or polypeptide of the present invention is preferably produced in purified form by conventional techniques. Typically, the protein or polypeptide of the present invention is secreted into the growth medium of recombinant *E. coli*. To isolate the protein, the *E. coli* host cell carrying a recombinant plasmid is propagated, homogenized, and the homogenate is centrifuged to remove bacterial debris. The supernatant is then subjected to sequential ammonium sulfate precipitation. The fraction containing the protein of the present invention is subjected to gel filtration in an appropriately sized dextran or polyacrylamide column to separate the proteins. If necessary, the protein fraction may be further purified by HPLC. Alternatively, the protein is purified by metal chelate affinity chromatography (Imai et al., "Expression and Purification of Functional Human 17α-hydroxylase/17,20-lyase ($P450_{c17}$) in *Escherichia coli*," Proc.

Natl. Acad. Sci. USA, 268:19681-19689 (1993); Kempf "Truncated Human P450 2D6: Expression in *Excherichia coli*, Ni^{2+}-chelate Affinity Purification, and Characterization of Solibility and Aggregation," Arch. Biochem. Biophys., 321:277-288 (1995), which are hereby incorporated by reference).

Mutations or variants of the above fusion protein are encompassed by the present invention.

Variants may be modified by, for example, the deletion or addition of amino acids that have minimal influence on the properties, secondary structure and hydropathic nature of the polypeptide. For example, a polypeptide may be conjugated to a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification, or identification of the polypeptide.

The DNA molecule encoding the cytochrome P450 polypeptide can be incorporated in cells using conventional recombinant DNA technology. Generally, this involves inserting the DNA molecule into an expression system to which the DNA molecule is heterologous (i.e. not normally present). The heterologous DNA molecule is inserted into the expression system or vector in proper sense orientation and correct reading frame. The vector contains the necessary elements for the transcription and translation of the inserted protein-coding sequences.

U.S. Patent No. 4,237,224 to Cohen and Boyer, which is hereby incorporated by reference, describes the production of expression systems in the form of recombinant plasmids using restriction enzyme cleavage and ligation with DNA ligase. These recombinant plasmids are then introduced by means of transformation and replicated in unicellular cultures including procaryotic organisms and eucaryotic cells grown in tissue culture.

Recombinant genes may also be introduced into viruses, such as vaccina virus. Recombinant viruses can be generated by transfection of plasmids into cells infected with virus.

Suitable vectors include, but are not limited to, the following viral vectors such as lambda vector system gt11, gt WES.tB, Charon 4, and plasmid vectors such as pBR322, pBR325, pACYC177, pACYC184, pUC8, pUC9, pUC18, pUC19,

pLG339, pR290, pKC37, pKC101, SV 40, pBluescript II SK +/- or KS +/- (see "Stratagene Cloning Systems" Catalog (1993) from Stratagene, La Jolla, Calif, which is hereby incorporated by reference), pQE, pIH821, pGEX, pET series (see F.W. Studier et. al., "Use of T7 RNA Polymerase to Direct Expression of Cloned Genes," Gene Expression Technology Vol. 185 (1990), which is hereby incorporated by reference), and any derivatives thereof. Recombinant molecules can be introduced into cells via transformation, particularly transduction, conjugation, mobilization, or electroporation. The DNA sequences are cloned into the vector using standard cloning procedures in the art, as described by Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Springs Laboratory, Cold Springs Harbor, New York (1982), which is hereby incorporated by reference.

A variety of host-vector systems may be utilized to express the protein-encoding sequence(s). Primarily, the vector system must be compatible with the host cell used. Host-vector systems include but are not limited to the following: bacteria transformed with bacteriophage DNA, plasmid DNA, or cosmid DNA; microorganisms such as yeast containing yeast vectors; mammalian cell systems infected with virus (e.g., vaccinia virus, adenovirus, etc.); insect cell systems infected with virus (e.g., baculovirus); and plant cells infected by bacteria. The expression elements of these vectors vary in their strength and specificities. Depending upon the host-vector system utilized, any one of a number of suitable transcription and translation elements can be used.

Different genetic signals and processing events control many levels of gene expression (e.g., DNA transcription and messenger RNA ("mRNA") translation).

Transcription of DNA is dependent upon the presence of a promotor which is a DNA sequence that directs the binding of RNA polymerase and thereby promotes mRNA synthesis. The DNA sequences of eucaryotic promotors differ from those of procaryotic promotors. Furthermore, eucaryotic promotors and accompanying genetic signals may not be recognized in or may not function in a procaryotic system, and, further, procaryotic promotors are not recognized and do not function in eucaryotic cells.

Similarly, translation of mRNA in procaryotes depends upon the presence of the proper procaryotic signals which differ from those of eucaryotes. Efficient translation of mRNA in procaryotes requires a ribosome binding site called the Shine-Dalgarno ("SD") sequence on the mRNA. This sequence is a short nucleotide sequence of mRNA that is located before the start codon, usually AUG, which encodes the amino-terminal methionine of the protein. The SD sequences are complementary to the 3'-end of the 16S rRNA (ribosomal RNA) and probably promote binding of mRNA to ribosomes by duplexing with the rRNA to allow correct positioning of the ribosome. For a review on maximizing gene expression see Roberts and Lauer, Methods in Enzymology, 68:473 (1979), which is hereby incorporated by reference.

Promotors vary in their "strength" (i.e. their ability to promote transcription). For the purposes of expressing a cloned gene, it is desirable to use strong promotors in order to obtain a high level of transcription and, hence, expression of the gene. Depending upon the host cell system utilized, any one of a number of suitable promotors may be used. For instance, when cloning in *E. coli*, its bacteriophages, or plasmids, promotors such as the T7 phage promoter, *lac* promotor, *trp* promotor, *rec*A promotor, ribosomal RNA promotor, the P_R and P_L promotors of coliphage lambda and others, including but not limited, to *lac*UV5, *omp*F, *bla*, *lpp*, and the like, may be used to direct high levels of transcription of adjacent DNA segments. Additionally, a hybrid *trp-lac*UV5 (*tac*) promotor or other *E. coli* promotors produced by recombinant DNA or other synthetic DNA techniques may be used to provide for transcription of the inserted gene.

Bacterial host cell strains and expression vectors may be chosen which inhibit the action of the promotor unless specifically induced. In certain operons, the addition of specific inducers is necessary for efficient transcription of the inserted DNA. For example, the *lac* operon is induced by the addition of lactose or IPTG (isopropylthio-beta-D-galactoside). A variety of other operons, such as *trp*, *pro*, etc., are under different controls.

Specific initiation signals are also required for efficient gene transcription and translation in procaryotic cells. These transcription and translation initiation signals may vary in "strength" as measured by the quantity of gene specific

messenger RNA and protein synthesized, respectively. The DNA expression vector, which contains a promotor, may also contain any combination of various "strong" transcription and/or translation initiation signals. For instance, efficient translation in *E. coli* requires a Shine-Dalgarno ("SD") sequence about 7-9 bases 5' to the initiation codon (ATG) to provide a ribosome binding site. Thus, any SD-ATG combination that can be utilized by host cell ribosomes may be employed. Such combinations include but are not limited to the SD-ATG combination from the *cro* gene or the *N* gene of coliphage lambda, or from the *E. coli* tryptophan E, D, C, B or A genes. Additionally, any SD-ATG combination produced by recombinant DNA or other techniques involving incorporation of synthetic nucleotides may be used.

Once the isolated DNA molecule encoding cytochrome P450 polypeptide has been cloned into an expression system, it is ready to be incorporated into a host cell. Such incorporation can be carried out by the various forms of transformation noted above, depending upon the vector/host cell system. Suitable host cells include, but are not limited to, bacteria, virus, yeast, mammalian cells, insect, and the like.

DNA molecules and nucleotide sequences which are derived from the disclosed DNA molecules as described above may also be defined as DNA sequences which hybridize under stringent conditions to the DNA sequences disclosed, or fragments thereof.

Suitable DNA molecules are those that hybridize to the chimeric DNA molecule under stringent conditions. An example of suitable high stringency conditions is when hybridization is carried out at 65°C for 20 hours in a medium containing 1M NaCl, 50 mM Tris-HCl, pH 7.4, 10 mM EDTA, 0.1% sodium dodecyl sulfate, 0.2% ficoll, 0.2% polyvinylpyrrolidone, 0.2% bovine serum albumin, 50 μm g/ml *E. coli* DNA.

In preferred embodiments of the present invention, stringent conditions may be defined as those under which DNA molecules with more than 25% sequence variation (also termed "mismatch") will not hybridize. Such conditions are referred to herein as conditions of 75% stringency (since hybridization will occur only between molecules with 75% homology or greater). In a more preferred embodiment, stringent conditions are those under which DNA molecules with more than 15%

mismatch will not hybridize (conditions of 85% stringency), and more preferably still, stringent conditions are those under which DNA sequences with more than 10% mismatch will not hybridize (conditions of 90% stringency). In a most preferred embodiment, stringent conditions are those under which DNA sequences with more than 6% mismatch will not hybridize (conditions of 94% stringency).

In yet another aspect of the present invention, the fusion protein can be applied to an environmental pollutant, such as an insecticide or other halogenated hydrocarbon spills, as part of a method of bioremediation. In fact, P450 enzymes can oxidize almost any compound that has a carbon-hydrogen bond and, thus, are useful for almost any environmental contaminant. Generally, microorganisms are extremely useful as agents for clean-up of environmental problems. Development of suitable microorganisms involves either selecting microorganisms with a bioremediation trait or by introducing a gene into microbes to engender them with that ability. By introducing the chimeric DNA molecule into an appropriate vector, it is possible to achieve bioremediation of environmental pollutants. Suitable vectors are non-pathogenic bacteria.

Another aspect of the present invention is using the fusion protein in a process of hydroxylating a compound to be oxidized. Typical compounds to be oxidized include hydrocarbons or any compound having a carbon-hydrogen bond. As discussed above, this involves contacting the compound to be oxidized with the fusion protein under conditions effective to hydroxylate the compound to be oxidized. The fusion protein can be provided by introducing the chimeric DNA molecule into an appropriate vector to express the fusion protein. Suitable vectors include pcW or pkk233-2.

Typicaly, hydroxylation occurs at from about 30 to about 50°C, with 37°C being preferred, with a potassium phosphate buffer and KCl (pH 7.4). The reaction can be monitored by the addition of dichloromethane and assaying by gas chromatography/mass spectrometry.

EXAMPLES

The following examples illustrate, but are not intended to limit, the present invention.

Example 1 - Construction of the Expression Plasmid for the Fusion Protein of $P450_{cam}$ and CYP2C9

CYP2C9 clone (pBP2C9) was obtained from the University of Washington, and $P450_{cam}$ (pBScam) was obtained from the University of Texas Southwestern Medical Center. Subcloning was performed in Epicurian Coli XL1-Blue MR supercompetent cells (Stratagene, LaJolla, CA). All modifications were introduced by PCR mutagenesis. Templates for PCR were pretreated by alkaline-denaturing method and, then, site-directed mutagenesis was performed by ExSite™PCR-Based Site-Directed Mutagenesis Kit (Stratagene, LaJolla, CA). Firstly, the *Nco* I restriction site was introduced in $P450_{cam}$ by primers 1 and 2 (the amino acids 216-218) and CYP2C9 by primers 3 and 4 (the amino acids 256-258). The starting position of the H-helix of CYP2C9 is aspartic acid 264. Since the homology model showed a conserved three-dimensional structure from the I-helix to the carboxy-terminus between $P450_{cam}$ and the CYP2C9 (Korzekwa et al., Pharacogenetics, 3:1-8 (1993), which is hereby incorporated by reference). The positions of amino acids were selected as a convenient conjunction. After digestion of *Xho* I ($P450_{cam}$) or *Eco* RI (CYP2C9), each plasmid was blunt-ended and, then, were digested by *Nco* I. The fragment of $P450_{cam}$ and CYP2C9 was ligated after the digestion by *Nco* I/*Xho* I or *Eco* RI. The ligated plasmid contained $P450_{cam}$, including the pBluescript vector, from the amino-terminus to the G-helix [1-216], and CYP2C9 from the H-helix to carboxy-terminus [Methionine 257 to C-terminus]. In addition, the sequence of junction [Ala-Met-Asp] was returned to the original sequence [Gly-Met-Asn] of $P450_{cam}$ or CYP2C9 by site-directed mutagenesis by primer 5 and 6. A $[His]_6$ affinity tag coding sequence was inserted at the 3'-terminus of CYP2C9 cDNA by primer 7 and 8. The sequences of the primers are:

primer 1 CCATGGACGCTATCAGCATCGTTGCCAAC (SEQ. ID. No. 7)

primer 2 CCGGCTTCTGCCTGCGTTGCTCGA (SEQ. ID. No. 8)

primer 3 CCATGGACAACCCTCAGGACTTTATTGAT (SEQ. ID. No. 9)

primer 4 CCATTGATTCTTGGTGTTCTTTTACT (SEQ. ID. No. 10)

primer 5 GCATGAACAACCCTCAGGACTTTATTGA (SEQ. ID. No. 11)

primer 6 CCGGCTTCTGCCTGCGTTGCTCG (SEQ. ID. No. 12)

primer 7 CATCACCATCACCATCACTGAAGAAGAGCAGATGGCCTGGC (SEQ. ID. No. 13)

primer 8 GACAGGAATGAAGCACAGCTGGTA (SEQ. ID. No. 14)

Example 2 - Expression of the Fusion Protein

A single ampicillin-resistant colony of DH5α cells transformed with plasmid DNA was grown overnight at 37°C in Luria-Bertani medium containing 100 μg ampicillin ml^{-1}. A 0.5-ml aliquot was used to innoculate 50 ml of Terrific broth ("TB") and cultured for 10 h. This aliquot of 25 ml was used to innoculate 500 ml of TB media. Incubation at 37°C was continued for 19 h. The TB media was supplemented with ampicillin (100 μg ml^{-1}), 0.2% glucose, 100 μM δ-aminolevulinic acid, vitamins (100^{-1}w/w, Basal Medium Eagle Vitamin Solution, Gibco BRL, Grand Island, NY), and trace elements (2 mM $MgSO_4.7H_2O$, 0.1 mM $CaCl_2$, 1.0 μM $FeSO_4$, metal solution1, 50 μM H_3BO_4, 0.2 μM $CoCl_2.6H_2O$, 1 mM $CuSO_4.5H_2O$, 1 mM $MnCl_2.4H_2O$, 1 nM Na_2MoO_4 and 2 mM $ZnCl_2$). The cells were harvested by centrifugation at 5,000 g and 4°C for 10 min. The pellet was stored at -80°C before use.

Example 3 - Construction of Expression Plasmid for Pd and PdR

Nde I restriction site was introduced at the site of the initiation codon of the Pd or PdR plasmids by the procedures similar to those described above. After digestion of Pd by *Sma* I and digestion of PdR by *Mlu* I followed by blunt-ending, each plasmid was digested by *Nde* I. Gel purified DNA was cloned into PET-15, an expression vector (Novagene, Madison, WI), after digestion by *Xho* I and blunt-ending. *E. coli* strain BL21(DE3) was transformed with pETPd or pETPdR.

Pd and PdR were expressed as follows. Icoculum cultures (25 ml) of *E. coli* BL21(DE3), transformed with pETPd or pETPdR were grown at 37°C in M9 minimum medium supplemented with 100 μg ampicillin ml^{-1}, 0.5% glucose, vitamins, and trace elements as mentioned above. A 25-ml aliquot was used to inoculate 500 ml of M9 minimum medium and the flask was shaken for 1 h at 37°C, at which

time 0.4 mM isopropyl β-D-thiogalactoside was added to induce the synthesis of T7 RNA polymerase. Incubation at 37°C was continued for 3 h.

Attempts to make a soluble chimeric construct were based on a homology model of CYP2C9. This model was produced with the program Modeller (Sali et al., 234:779-815 (1993), which is hereby incorporated by reference), and used the coordinates of $P450_{cam}$, $P450_{BM3}$ and $P450_{eryF}$. The resulting homology model indicated that replacing all amino acids prior to the random coil between the G- and H-helix (using $P450_{cam}$ structural nomenclature) with bacterial amino acids may provide a soluble bacterial/mammalian chimera. This coil was chosen, because it was believed that amino-terminus and possibly the distal face of the protein (comprised of amino acids prior to the coil) were involved in membrane interactions. Furthermore, since the sequence alignments are based on very low sequence identity, it was believed that by choosing an area for fusion with no secondary structure chances of producing a folded protein would increase.

A chimera was based on the homology model to contain $P450_{cam}$ from the amino-terminus to the G-helix [1-216] and CYP2C9 from before the putative H-helix to carboxy-terminus [Methionine 257 to C-terminus] (Figures 1(A) and (B)). According to the nomenclature of Gotoh, O. J. Biol Chem., 267:83-90 (1992), which is hereby incorporated by reference, the active site would be composed of SRS (substrate recognition site)1-3 from $P450_{cam}$ and SRS4-6 from P450 2C9. All modifications were introduced by PCR-mutagenesis (Dorrell et al., "Improved Efficiency of Inverse PCR Mutagenesis," BioTechniques, 21:604-608 (1996), which is hereby incorporated by reference). $A[His]_6$ affinity tag coding sequence was inserted at the 3'-terminus of P450 2C9 cDNA to allow protein purification by metal chelate affinity chromatograph. (Imai et al., "Expression and Purification of Functional Human 17α-hydroxylase/17,20-lyase ($P450_{c17}$) in *Escherichia coli*," Proc. Natl. Acad. Sci. USA, 268:19681-19689 (1993); Kempf "Truncated Human P450 2D6: Expression in *Excherichia coli*, Ni^{2+}-chelate Affinity Purification, and Characterization of Solibility and Aggregation," Arch. Biochem. Biophys., 321:277-288 (1995), which are hereby incorporated by reference). The protein was expressed in *E. coli* with the pBluescript vector. This preparation yielded 260 nmol/liter of Terrific broth medium after 29 h of culture at 37°C. (Peterson et al., "Putidaredoxin

Reductase and Puridaredoxin: Cloning, Sequence, and Heterologous Expression of the Proteins," J. Biol. Chem., 265:6066-6073 (1990), which is hereby incorporated by reference). Expression levels of the wild type $P450_{cam}$ was 600-1000 nmoles/liter under similar conditions. After treatment with lysozyme and sonication of the cell pellet, the cell lysate was centrifuged at 105,000g and the supernatant was applied to a Ni-NTA agarose and hydroxylapatite columns (Imai et al., "Expression and Purification of Functional Human 17α-hydroxylase/17,20-lyase (P45017) in *Escherichia coli*," Proc. Natl. Acad. Sci. USA, 268:19681-19689 (1993), which is hereby incorporated by reference). The purified chimera showed a CO-reduced difference spectrum at 448 nm (Fig. 2A) (Omura et al., "The Carbon Monoxide-Binding Pigment of Liver Microsomes I Evidence for its Hemeprotein Nature," J. Biol. Chem., 239:2370-2378 (1964), which is hereby incorporated by reference), and showed two major bands on SDS-polyacrylamide gel electrophoresis (Fig. 2B) (Laemmli, U.K., "Cleavage of Structural Protein During the Assembly of the Head of Bacteriophage," Nature, 227:680-685 (1970), which is hereby incorporated by reference). Similar bands are observed from purified wild-type P450cam with a $[His]_6$ tag coding sequence. The lower molecule weight band is presently unidentified. The resulting purified protein showed an approximae molecular weight of 51 kDa as judged by SDS-polyacrylamide gel electrophoresis, consistent with the molecular weight expected for the chimera (Figure 2B).

The resulting pruified protein showed a reduced CO difference spectrum at 450 nm (Figure 2A). These data are consistent with a folded P450 protein having a functional active site. The observation that a functional chimera of P450 2C9 and $P450_{cam}$, which have only 15% primary sequence homology, can still bind CO provides strong evidence for a conserved three-dimensional structure between $P450_{cam}$ and CYP2 family. The fact that the resulting enzyme is soluble, while mammalian enzymes with the amino terminus removed are not, indicates that other regions near the amino terminus may also be important for membrane interactions. (Lemos-Chiarandine et al., J. Cell Biol., 104:209-219 (1987); Vergeres et al., Biochemistry, 28:3650-3655 (1989); Wachenfeldt et al., Arch. Biochem. Biophys., 339:107-114 (1997), which are hereby incorporated by reference.)

Since CO binding spectra is only an indirect measure of whether the chimeric protein has folded, circular dichroism studies were performed to explore the secondary structure of the bacterial/mammalian chimera. (Pfeil et al., Biochemistry, 32:8856-62 (1993), which is hereby incorporated by reference). The spectrum of the chimera showed a typical helix structure (data not shown). The predicted secondary structure based on these studies are presented in Table 1.

Table 1

	Fraction	Chimera Ratio	$P450_{cam}$ Ratio
Helix:	0.2	35.5	28.8
Beta:	0.0	5.4	18.0
Turn:	0.2	23.2	20.8
Random:	0.2	35.8	32.4
Total	0.7	100.0	100.0

The predicted amount of α-helix and β-sheet secondary structure were similar between the chimera and $P450_{cam}$ wild type. Thus, the circular dichroism studies confirm that the chimera is folded and has similar secondary structural features as the bacterial $P450_{cam}$.

Next, the ability of the fusion protein to oxidize a common P450 substrate was determined. The bacterial and mammalian enzymes both require an electron tranfer protein to reduce molecular oxygen to an active monooxygen oxidant. However, the bacterial and mammaliam enzyme use different unrelated electron transfer proteins. To determine if the bacterial electron transfer proteins could function as an electron donor, putidaredoxin and putidaredoxin reductase were purified after subcloning their cDNAs to pET vector the *T7lac* promoter and $[His]_6$ taggled sequence. This bacterial electron transfer system could support the oxidation of 4-chlorotoluene to 4-chlorobenzyl alcohol by the fusion protein. The hydroxylation occured at 37°C being preferred. 50 mM potassium phosphate buffer was utilized with 200 MM KCl, (pH 7.4). Each reaction contained 500 μM 4-chlorotoluene, between .4 and 1 nmole of P450, 3 μM putidaredoxin, 1.5 μM

putidaredoxin reductase, and 300 μM NADH. The reaction was stopped by the addition of 4 ml of dichloromethane and assayed by gas chromatography/mass spectrometry. Experiments to determine if the mammalian P450 reductase can support the same oxidation are underway.

Detection of the catalyic activity toward 4-chlorotoluene indicate that the fusion protein can function as an active P450 enzyme (Table 1). As compared with the turnover number from the wild type $P450_{cam}$, the chimera shows approximately 3 times the activity towards 4-chlorotoluene. This means a potential for making soluble P450 that can perform stereospecific synthesis.

This approach could have a number of applications. 1) From other homology models of mammalian P450 enzymes it is apparent that this method may prove to be a general method for constructed soluble P450 enzymes with mammalian active site characteristics. These enzymes should be more adaptable to uses in benign synthesis and bioremediation than the more restrictive bacterial enzymes and easier to work with then the membrane bound mammalian enzymes. 2) Selectively replacing amino acid segments in the amino terminus with the mammalian amino acids may prove to be a valuable method of determining important membrane association sites. 3) Since the enzyme is soluble, it could prove a method for obtaining structural information. In particular it should be amiable to Xray crystallography. 4) Since the enzyme is part mammalian and part bacterial, it can be used to determine the features that confer specific interactions with the different reductases system that are used by the bacterial and mammalian proteins.

Although the invention has been described in detail for the purpose of illustration, it is understood that such detail is solely for that purpose, and variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention which is defined by the following claims.

WHAT IS CLAIMED:

1. A chimeric DNA molecule comprising:

a first DNA molecule encoding a portion of a full length bacterial P450 protein;

a second DNA molecule fused to the first DNA molecule and encoding a portion of a full length mammalian P450 protein, wherein the chimeric DNA molecule encodes a fusion protein which is active and soluble in aqueous liquid.

2. A chimeric DNA molecule according to claim 1, wherein the first and second DNA molecules are fused together at a location where the encoded fusion protein lacks secondary structure.

3. A chimeric DNA molecule according to claim 1, wherein the chimeric DNA molecule is prepared from a DNA molecule encoding a full length mammalian P450 protein where a portion of the DNA molecule encoding a full length mammalian P450 protein is replaced with a DNA molecule encoding a homologous portion of a full length bacterial P450 protein.

4. A chimeric DNA molecule according to claim 3, wherein all amino acids prior to a random coil between G- and H-helices in the full length mammalian P450 protein are replaced with a homologous portion of the full length bacterial P450 protein.

5. A chimeric DNA molecule according to claim 3, wherein the chimeric DNA molecule comprises about 50 percent of the DNA molecule encoding the full length mammalian P450 protein and about 50 percent of the DNA molecule encoding the full length bacterial P450 protein.

6. A chimeric DNA molecule according to claim 1, wherein the second DNA molecule encodes a portion of CYP2C9.

7. A chimeric DNA molecule according to claim 1, wherein the first DNA molecule encodes a portion of $P450_{cam}$.

8. A chimeric DNA molecule according to claim 1, wherein the chimeric DNA molecule has a heme ligand positioned in a relative orientation to an I-helix and a fifth cysteine ligand similar to that of the heme ligand in a full length mammalian P450 protein.

9. A chimeric DNA molecule according to claim 1, wherein the chimeric DNA molecule encodes an amino acid sequence of SEQ. ID. No. 2.

10. A chimeric DNA molecule according to claim 9, wherein the chimeric DNA molecule has a nucleotide sequence of SEQ. ID. No. 1.

11. A DNA expression system transformed with the chimeric DNA molecule of claim 1.

12. A DNA expression system according to claim 11, wherein the chimeric DNA molecule is positioned in the expression system in proper sense orientation and correct reading frame.

13. A DNA expression system according to claim 11, wherein the first and second DNA molecules are fused together at a location where the encoded fusion protein lacks secondary structure.

14. A host cell transformed with the chimeric DNA molecule of claim 1.

15. A host cell according to claim 14, wherein the host cell is selected from the group consisting of plant cells, mammalian cells, insect cells, and bacterial cells.

16. A fusion protein comprising:

a portion of a bacterial P450 protein and

a portion of a mammalian P450 protein fused to the portion of a bacterial P450 protein, wherein the fusion protein is active and soluble in aqueous liquid.

17. A fusion protein according to claim 16, wherein the portion of a mammalian P450 protein and the portion of a bacterial P450 protein are fused where the encoded fusion protein lacks secondary structure.

18. A fusion protein according to claim 16, wherein the fusion protein is prepared from a full length mammalian P450 protein where a portion of the full length mammalian P450 protein is replaced with a homologous portion of a full length bacterial P450 protein.

19. A fusion protein according to claim 18, wherein all amino acids prior to a random coil between G- and H-helices in the full length mammalian P450 protein are replaced with a homologous portion of the full length bacterial P450 protein.

20. A fusion protein according to claim 18, wherein the fusion protein comprises about 50 percent of the full length mammalian P450 protein and about 50 percent of the full length bacterial P450 protein.

21. A fusion protein according to claim 16, wherein the mammalian P450 protein is CYP2C9.

22. A fusion protein according to claim 16, wherein the bacterial P450 protein is $P450_{cam}$.

23. A fusion protein according to claim 16, wherein the fusion protein has a heme ligand positioned in a relative orientation to an I-helix and a fifth cysteine ligand similar to that of the heme ligand in a full length mammalian P450 protein.

24. A fusion protein according to claim 16, wherein the fusion protein has an amino acid sequence of SEQ. ID. No. 2.

25. A method of hydroxylating a compound to be oxidized comprising:

contacting the compound to be oxidized with the fusion protein according to claim 16 under conditions effective to hydroxylate the compound to be oxidized.

26. A method according to claim 25, wherein the portion of the mammalian P450 protein and the portion of the bacterial P450 protein are fused where the encoded fusion protein lacks secondary structure.

27. A method according to claim 25, wherein the fusion protein is prepared from a full length mammalian P450 protein where a portion of the full length mammalian P450 protein is replaced with a homologous portion of a full length bacterial P450 protein.

28. A method according to claim 27, wherein all amino acids prior to a random coil between G- and H-helices in the full length mammalian P450 protein are replaced with a homologous portion of the full length bacterial P450 protein.

29. A method according to claim 27, wherein the fusion protein comprises about 50 percent of the full length mammalian P450 protein and about 50 percent of the full length bacterial P450 protein.

30. A method according to claim 25, wherein the fusion protein is provided by providing a vector comprising a chimeric DNA molecule comprising:

a first DNA molecule encoding a portion of a full length bacterial P450 protein;

a second DNA molecule fused to the first DNA molecule and encoding a portion of a full length mammalian P450 protein, wherein the chimeric DNA molecule encodes the fusion protein.

31. A method according to claim 30, wherein the first and second DNA molecules are fused together at a location where the encoded fusion protein lacks secondary structure.

32. A method according to claim 30, wherein the chimeric DNA molecule is prepared from a DNA molecule encoding a full length mammalian P450

protein where a portion of the DNA molecule encoding a full length mammalian P450 protein is replaced with a DNA molecule encoding a homologous portion of a full length bacterial P450 protein.

33. A method according to claim 32, wherein all amino acids prior to a random coil between G- and H-helices in the full length mammalian P450 protein are replaced with a homologous portion of the full length bacterial P450 protein.

34. A method according to claim 32, wherein the chimeric DNA molecule comprises about 50 percent of the DNA molecule encoding the full length mammalian P450 protein and about 50 percent of the DNA molecule encoding the full length bacterial P450 protein.

35. A method of bioremediation of an environmental pollutant comprising:

contacting the environmental pollutant with a fusion protein according to claim 16 under conditions effective to effect bioremediation.

36. A method according to claim 35, wherein the portion of the mammalian P450 protein and the portion of the bacterial P450 protein are fused where the encoded fusion protein lacks secondary structure.

37. A method according to claim 35, wherein the fusion protein is prepared from a full length mammalian P450 protein where a portion of the full length mammalian P450 protein is replaced with a homologous portion of a full length bacterial P450 protein.

38. A method according to claim 37, wherein all amino acids prior to a random coil between G- and H-helices in the full length mammalian P450 protein are replaced with a homologous portion of the full length bacterial P450 protein.

39. A method according to claim 37, wherein the fusion protein comprises about 50 percent of the full length mammalian P450 protein and about 50 percent of the full length bacterial P450 protein.

40. A method according to claim 35, wherein the fusion protein is provided by providing a vector comprising a chimeric DNA molecule comprising:

a first DNA molecule encoding a portion of a full length bacterial P450 protein;

a second DNA molecule fused to the first DNA molecule and encoding a portion of a full length mammalian P450 protein, wherein the chimeric DNA molecule encodes the fusion protein.

41. A method according to claim 40, wherein the first and second DNA molecules are fused together at a location where the encoded fusion protein lacks secondary structure.

42. A method according to claim 40, wherein the chimeric DNA molecule is prepared from a DNA molecule encoding a full length mammalian P450 protein where a portion of the DNA molecule encoding a full length mammalian P450 protein is replaced with a DNA molecule encoding a homologous portion of a full length bacterial P450 protein.

43. A method according to claim 42, wherein all amino acids prior to a random coil between G- and H-helices in the full length mammalian P450 protein are replaced with a homologous portion of the full length bacterial P450 protein.

44. A method according to claim 42, wherein the chimeric DNA molecule comprises about 50 percent of the DNA molecule encoding the full length mammalian P450 protein and about 50 percent of the DNA molecule encoding the full length bacterial P450 protein.

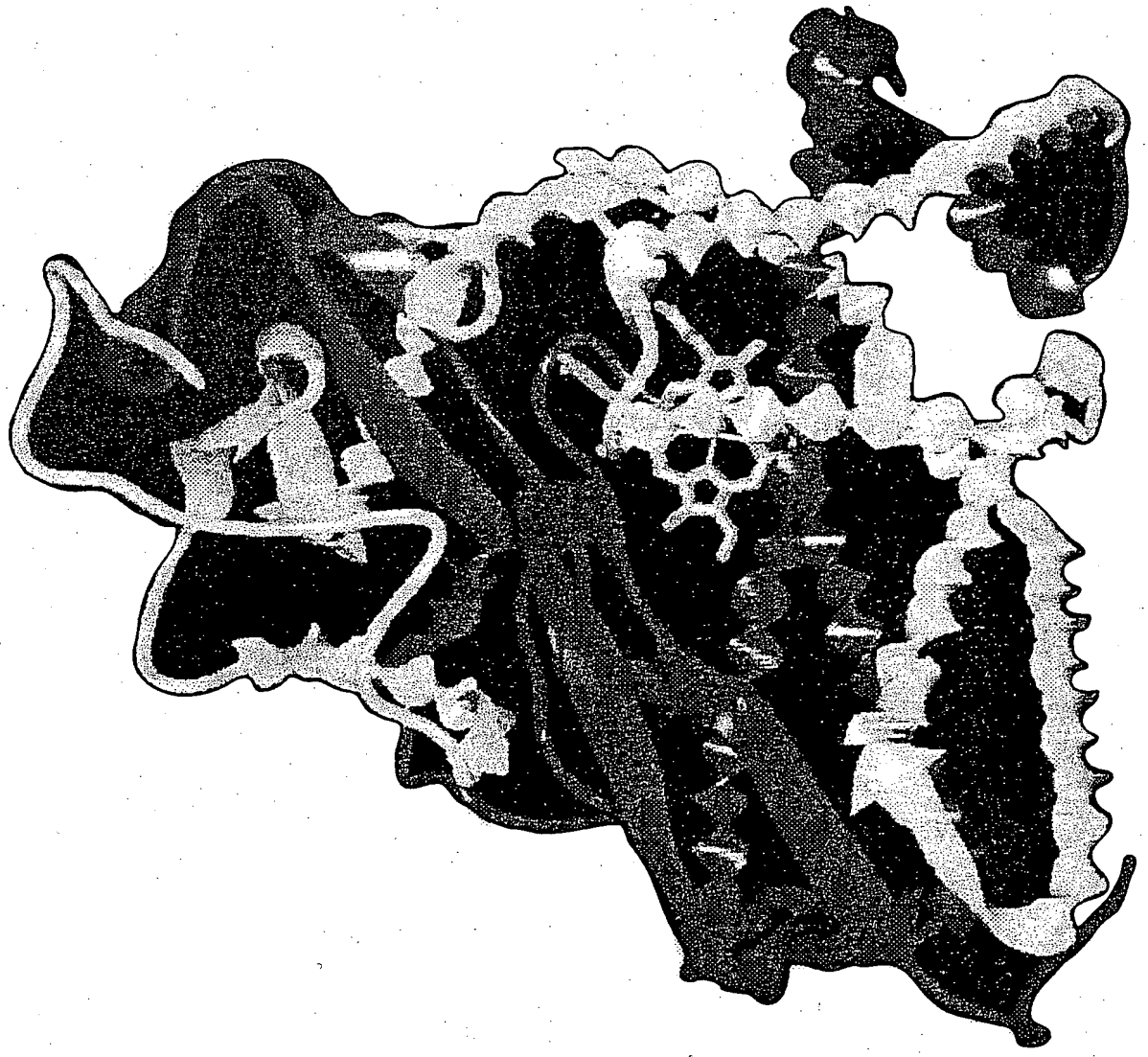

FIG. 1A

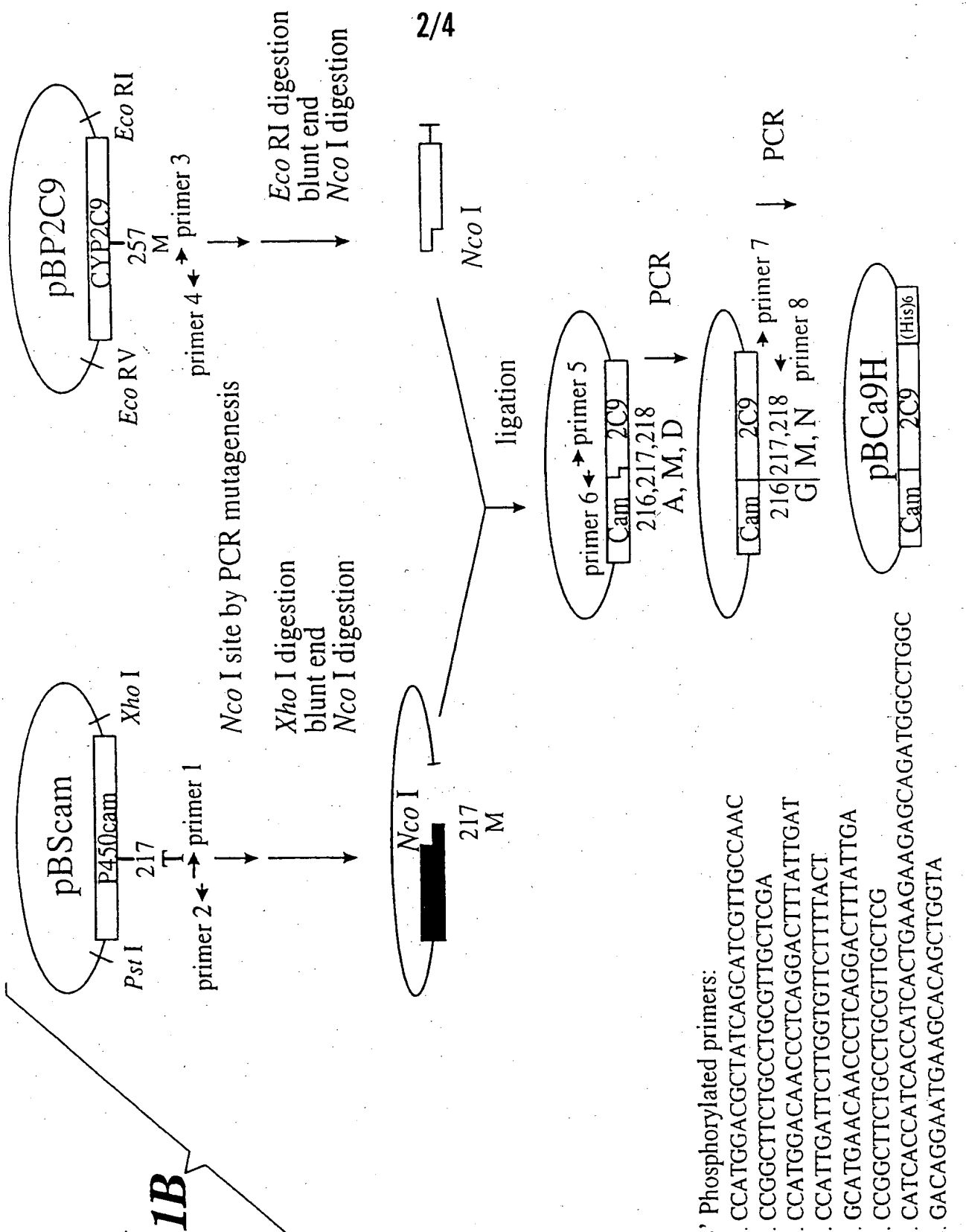

FIG. 1B

SCAN SPEED: 500 nM/min

PEAK PICK		POINT PICK	
λ	Abs	λ	Abs
495.0	-0.05	420.0	-0.58
446.5	1.234	448.0	1.221
		450.0	1.144
		480.0	0.02
		495.0	-0.05

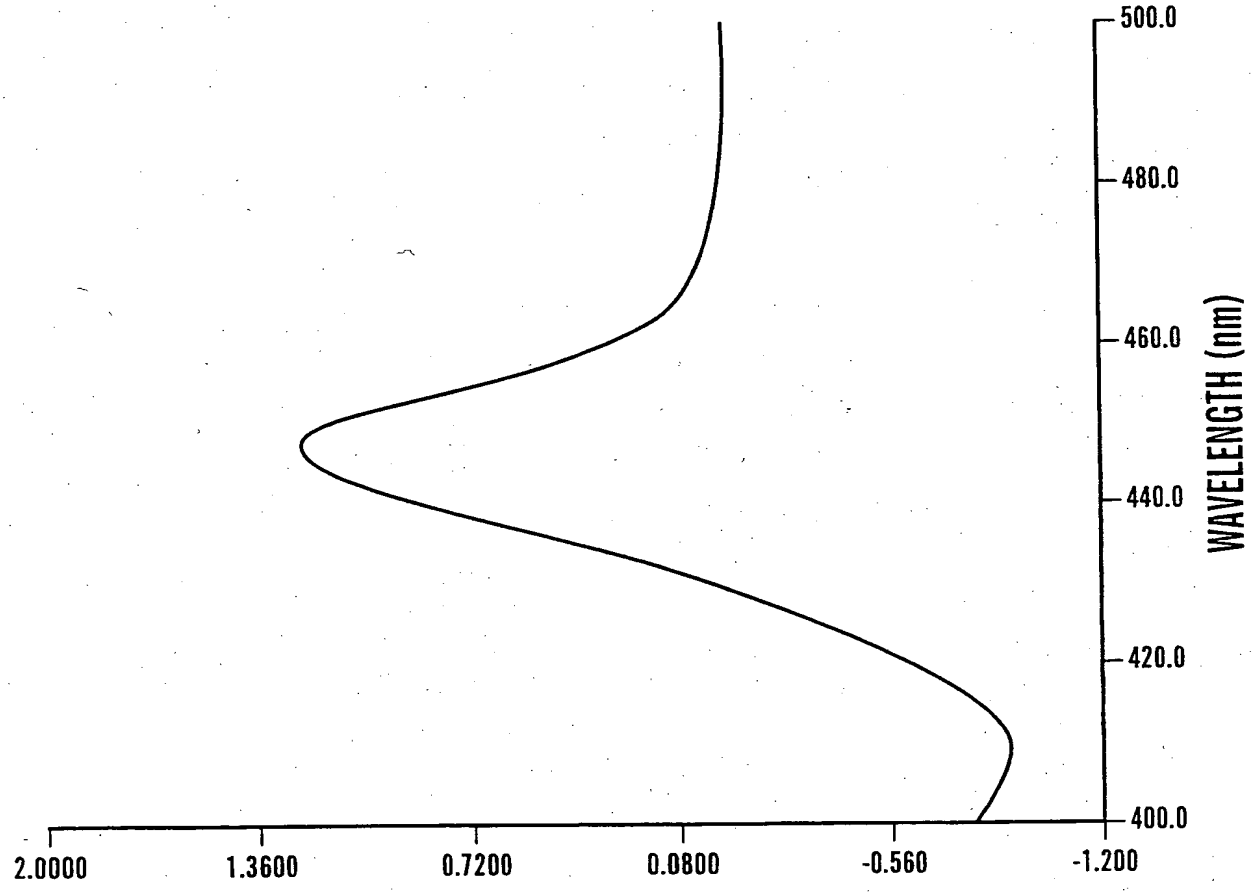

FIG. 2A

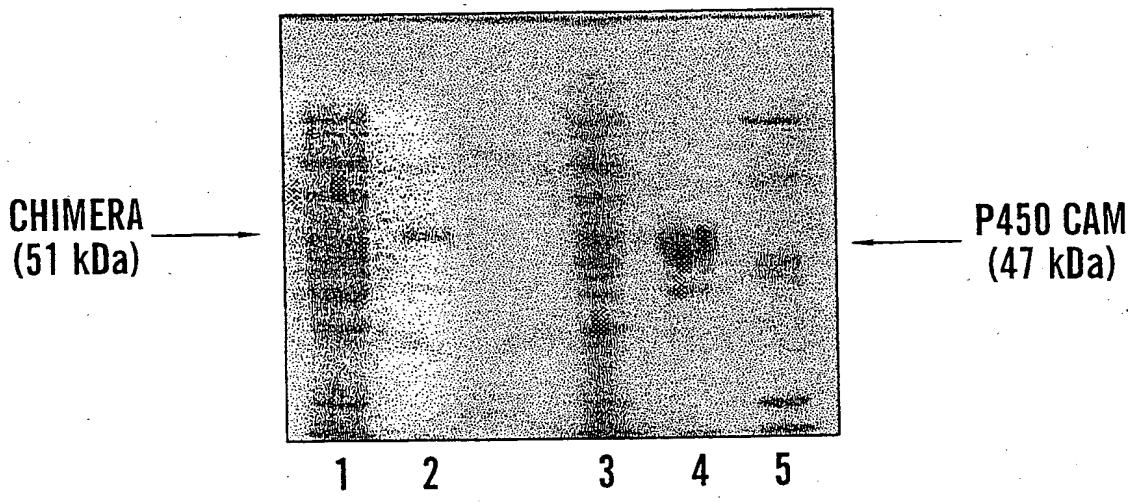

FIG. 2B

SEQUENCE LISTING

<110> University of Rochester

<120> FUNCTIONAL BACTERIAL/MAMMALIAN CYTOCHROME P450 CHIMERA

<130> 176/60232

<140>
<141>

<150> 60/056,754
<151> 1997-08-20

<160> 14

<170> PatentIn Ver. 2.0

<210> 1
<211> 1356
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: chimera of bacterial and mammalian

<400> 1
atgacgactg aaaccataca aagcaacgcc aatcttgccc ctctgccacc ccatgtgcca 60
gagcacctgg tattcgactt cgacatgtac aatccgtcga atctgtctgc cggcgtgcag 120
gaggcctggg cagttctgca agaatcaaac gtaccggatc tggtgtggac tcgctgcaac 180
ggcggacact ggatcgccac tcgcggccaa ctgatccgtg aggcctatga agattaccgc 240
cacttttcca gcgagtgccc gttcatccct cgtgaagccg gcgaagccta cgacttcatt 300
cccacctcga tggatccgcc cgagcagcgc cagtttcgtg cgctggccaa ccaagtggtt 360
ggcatgccgg tggtggataa gctggagaac cggatccagg agctggcctg ctcgctgatc 420
gagagcctgc gcccgcaagg acagtgcaac ttcaccgagg actacgccga acccttcccg 480
atacgcatct tcatgctgct cgcaggtcta ccggaagaag atatcccgca cttgaaatac 540
ctaacggatc agatgacccg tccggatggc agcatgacct tcgcagaggc caaggaggcg 600
ctctacgact atctgatacc gatcatcgag caacgcaggc agaagccggg aatgaacaac 660
cctcaggact ttattgattg cttcctgatg aaaatggaga aggaaaagca caaccaacca 720
tctgaattta ctattgaaag cttggaaaac actgcagttg acttgtttgg agctgggaca 780
gagacgacaa gcacaaccct gagatatgct ctccttctcc tgctgaagca cccagaggtc 840
acagctaaag tccaggaaga gattgaacgt gtgattggca gaaaccggag cccctgcatg 900
caagacagga gccacatgcc ctacacagat gctgtggtgc acgaggtcca gagatacatt 960
gaccttctcc ccaccagcct gccccatgca gtgacctgtg acattaaatt cagaaactat 1020
ctcattccca agggcacaac catattaatt tccctgactt ctgtgctaca tgacaacaaa 1080
gaatttccca acccagagat gtttgaccct catcactttc tggatgaagg tggcaatttt 1140
aagaaaagta aatacttcat gcctttctca gcaggaaaac ggatttgtgt gggagaagcc 1200

ctggccggca tggagctgtt tttattcctg acctccattt tacagaactt taacctgaaa 1260
tctctggttg acccaaagaa ccttgacacc actccagttg tcaatggatt tgcctctgtg 1320
ccgcccttct accagctgtg cttcattcct gtctga 1356

<210> 2
<211> 446
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: fusion protein

<400> 2

Asn Leu Ala Pro Leu Pro Pro His Val Pro Glu His Leu Val Phe Asp
1 5 10 15

Phe Asp Met Tyr Asn Pro Ser Asn Leu Ser Ala Gly Val Gln Glu Ala
20 25 30

Trp Ala Val Leu Gln Glu Ser Asn Val Pro Asp Leu Val Trp Thr Arg
35 40 45

Cys Asn Gly Gly His Trp Ile Ala Thr Arg Gly Gln Leu Ile Arg Glu
50 55 60

Ala Tyr Glu Asp Tyr Arg His Phe Ser Ser Glu Cys Pro Phe Ile Pro
65 70 75 80

Arg Glu Ala Gly Glu Ala Tyr Asp Phe Ile Pro Thr Ser Met Asp Pro
85 90 95

Pro Glu Gln Arg Gln Phe Arg Ala Leu Ala Asn Gln Val Val Gly Met
100 105 110

Pro Val Val Asp Lys Leu Glu Asn Arg Ile Gln Glu Leu Ala Cys Ser
115 120 125

Leu Ile Glu Ser Leu Arg Pro Gln Gly Gln Cys Asn Phe Thr Glu Asp
130 135 140

Tyr Ala Glu Pro Phe Pro Ile Arg Ile Phe Met Leu Leu Ala Gly Leu
145 150 155 160

Pro Glu Glu Asp Ile Pro His Leu Lys Tyr Leu Thr Asp Gln Met Thr
165 170 175

Arg Pro Asp Gly Ser Met Thr Phe Ala Glu Ala Lys Glu Ala Leu Tyr
180 185 190

Asp Tyr Leu Ile Pro Ile Ile Glu Gln Arg Arg Gln Lys Pro Gly Asn
195 200 205

Asn Pro Gln Asp Phe Ile Asp Cys Phe Leu Met Lys Met Glu Lys Glu
210 215 220

Lys His Asn Gln Pro Ser Glu Phe Thr Ile Glu Ser Leu Glu Asn Thr
225 230 235 240

Ala Val Asp Leu Phe Gly Ala Gly Thr Glu Thr Thr Ser Thr Thr Leu
245 250 255

Arg Tyr Ala Leu Leu Leu Leu Leu Lys His Pro Glu Val Thr Ala Lys
260 265 270

Val Gln Glu Glu Ile Glu Arg Val Ile Gly Arg Asn Arg Ser Pro Cys
275 280 285

Met Gln Asp Arg Ser His Met Pro Tyr Thr Asp Ala Val Val His Glu
290 295 300

Val Gln Arg Tyr Ile Asp Leu Leu Pro Thr Ser Leu Pro His Ala Val
305 310 315 320

Thr Cys Asp Ile Lys Phe Arg Asn Tyr Leu Ile Pro Lys Gly Thr Thr
325 330 335

Ile Leu Ile Ser Leu Thr Ser Val Leu His Asp Asn Lys Glu Phe Pro
340 345 350

Asn Pro Glu Met Phe Asp Pro His His Phe Leu Asp Glu Gly Gly Asn
355 360 365

Phe Lys Lys Ser Lys Tyr Phe Met Pro Phe Ser Ala Gly Lys Arg Ile
370 375 380

Cys Val Gly Glu Ala Leu Ala Gly Met Glu Leu Phe Leu Phe Leu Thr
385 390 395 400

Ser Ile Leu Gln Asn Phe Asn Leu Lys Ser Leu Val Asp Pro Lys Asn
405 410 415

Leu Asp Thr Thr Pro Val Val Asn Gly Phe Ala Ser Val Pro Pro Phe
420 425 430

Tyr Gln Leu Cys Phe Ile Pro Val His His His His His His
435 440 445

<210> 3
<211> 490
<212> PRT
<213> mammalian

<400> 3

```
Met Asp Ser Leu Val Val Leu Val Leu Cys Leu Ser Cys Leu Leu Leu
  1               5                  10                  15

Leu Ser Leu Trp Arg Gln Ser Ser Gly Arg Gly Lys Leu Pro Pro Gly
             20                  25                  30

Pro Thr Pro Leu Pro Val Ile Gly Asn Ile Leu Gln Ile Gly Ile Lys
         35                  40                  45

Asp Ile Ser Lys Ser Leu Thr Asn Leu Ser Lys Val Tyr Gly Pro Val
     50                  55                  60

Phe Thr Leu Tyr Phe Gly Leu Lys Pro Ile Val Val Leu His Gly Tyr
 65                  70                  75                  80

Glu Ala Val Lys Glu Ala Leu Ile Asp Leu Gly Glu Glu Phe Ser Gly
                 85                  90                  95

Arg Gly Ile Phe Pro Leu Ala Glu Arg Ala Asn Arg Gly Phe Gly Ile
            100                 105                 110

Val Phe Ser Asn Gly Lys Lys Trp Lys Glu Ile Arg Arg Phe Ser Leu
        115                 120                 125

Met Thr Leu Arg Asn Phe Gly Met Gly Lys Arg Ser Ile Glu Asp Arg
    130                 135                 140

Val Gln Glu Glu Ala Arg Cys Leu Val Glu Glu Leu Arg Lys Thr Lys
145                 150                 155                 160

Ala Ser Pro Cys Asp Pro Thr Phe Ile Leu Gly Cys Ala Pro Cys Asn
                165                 170                 175

Val Ile Cys Ser Ile Ile Phe His Lys Arg Phe Asp Tyr Lys Asp Gln
            180                 185                 190

Gln Phe Leu Asn Leu Met Glu Lys Leu Asn Glu Asn Ile Lys Ile Leu
        195                 200                 205

Ser Ser Pro Trp Ile Gln Ile Cys Asn Asn Phe Ser Pro Ile Ile Asp
```

210 215 220

Tyr Phe Pro Gly Thr His Asn Lys Leu Leu Lys Asn Val Ala Phe Met
225 230 235 240

Lys Ser Tyr Ile Leu Glu Lys Val Lys Glu His Gln Glu Ser Met Asp
245 250 255

Met Asn Asn Pro Gln Asp Phe Ile Asp Cys Phe Leu Met Lys Met Glu
260 265 270

Lys Glu Lys His Asn Gln Pro Ser Glu Phe Thr Ile Glu Ser Leu Glu
275 280 285

Asn Thr Ala Val Asp Leu Phe Gly Ala Gly Thr Glu Thr Thr Ser Thr
290 295 300

Thr Leu Arg Tyr Ala Leu Leu Leu Leu Leu Lys His Pro Glu Val Thr
305 310 315 320

Ala Lys Val Gln Glu Glu Ile Glu Arg Val Ile Gly Arg Asn Arg Ser
325 330 335

Pro Cys Met Gln Asp Arg Ser His Met Pro Tyr Thr Asp Ala Val Val
340 345 350

His Glu Val Gln Arg Tyr Ile Asp Leu Leu Pro Thr Ser Leu Pro His
355 360 365

Ala Val Thr Cys Asp Ile Lys Phe Arg Asn Tyr Leu Ile Pro Lys Gly
370 375 380

Thr Thr Ile Leu Ile Ser Leu Thr Ser Val Leu His Asp Asn Lys Glu
385 390 395 400

Phe Pro Asn Pro Glu Met Phe Asp Pro His His Phe Leu Asp Glu Gly
405 410 415

Gly Asn Phe Lys Lys Ser Lys Tyr Phe Met Pro Phe Ser Ala Gly Lys
420 425 430

Arg Ile Cys Val Gly Glu Ala Leu Ala Gly Met Glu Leu Phe Leu Phe
435 440 445

Leu Thr Ser Ile Leu Gln Asn Phe Asn Leu Lys Ser Leu Val Asp Pro
450 455 460

Lys Asn Leu Asp Thr Thr Pro Val Val Asn Gly Phe Ala Ser Val Pro

465 470 475 480

Pro Phe Tyr Gln Leu Cys Phe Ile Pro Val
485 490

<210> 4
<211> 1845
<212> DNA
<213> mammalian

<400> 4
gaaggcttca atggattctc ttgtggtcct tgtgctctgt ctctcatgtt tgcttctcct 60
ttcactctgg agacagagct ctgggagagg aaaactccct cctggcccca ctcctctccc 120
agtgattgga aatatcctac agataggtat taaggacatc agcaaatcct taaccaatct 180
ctcaaaggtc tatggccctg tgttcactct gtattttggc ctgaaaccca tagtggtgct 240
gcatggatat gaagcagtga aggaagccct gattgatctt ggagaggagt tttctggaag 300
aggcatttc ccactggctg aaagagctaa cagaggattt ggaattgttt tcagcaatgg 360
aaagaaatgg aaggagatcc ggcgtttctc cctcatgacg ctgcggaatt ttgggatggg 420
gaagaggagc attgaggacc gtgttcaaga ggaagcccgc tgccttgtgg aggagttgag 480
aaaaaccaag gcctcaccct gtgatcccac tttcatcctg ggctgtgctc cctgcaatgt 540
gatctgctcc attattttcc ataaacgttt tgattataaa gatcagcaat ttcttaactt 600
aatggaaaag ttgaatgaaa acatcaagat tttgagcagc ccctggatcc agatctgcaa 660
taatttttct cctatcattg attacttccc gggaactcac aacaaattac ttaaaaacgt 720
tgcttttatg aaaagttata ttttggaaaa agtaaaagaa caccaagaat caatggacat 780
gaacaaccct caggacttta ttgattgctt cctgatgaaa atggagaagg aaaagcacaa 840
ccaaccatct gaatttacta ttgaaagctt ggaaaacact gcagttgact tgtttggagc 900
tgggacagag acgacaagca caaccctgag atatgctctc cttctcctgc tgaagcaccc 960
agaggtcaca gctaaagtcc aggaagagat tgaacgtgtg attggcagaa accggagccc 1020
ctgcatgcaa gacaggagcc acatgcccta cacagatgct gtggtgcacg aggtccagag 1080
atacattgac cttctcccca ccagcctgcc ccatgcagtg acctgtgaca ttaaattcag 1140
aaactatctc attcccaagg gcacaaccat attaatttcc ctgacttctg tgctacatga 1200
caacaaagaa tttcccaacc cagagatgtt tgaccctcat cactttctgg atgaaggtgg 1260
caattttaag aaaagtaaat acttcatgcc tttctcagca ggaaaacgga tttgtgtggg 1320
agaagccctg gccggcatgg agctgttttt attcctgacc tccattttac agaactttaa 1380
cctgaaatct ctggttgacc caaagaacct tgacaccact ccagttgtca atggatttgc 1440
ctctgtgccg cccttctacc agctgtgctt cattcctgtc tgaagaagag cagatggcct 1500
ggctgctgct gtgcagtccc tgcagctctc tttcctctgg ggcattatcc atctttcact 1560
atctgtaatg ccttttctca cctgtcatct cacattttcc cttccctgaa gatctagtga 1620
acattcgacc tccattacgg agagtttcct atgtttcact gtgcaaatat atctgctatt 1680
ctccatactc tgtaacagtt gcattgactg tcacataatg ctcatactta tctaatgttg 1740
agttattaat atgttattat taaatagaga aatatgattt gtgtattata attcaaaggc 1800
atttcttttc tgcatgttct aaataaaaag cattattatt tgctg 1845

<210> 5
<211> 405
<212> PRT
<213> Pseudomonas putida

<400> 5

```
Asn Leu Ala Pro Leu Pro Pro His Val Pro Glu His Leu Val Phe Asp
 1               5                  10                  15

Phe Asp Met Tyr Asn Pro Ser Asn Leu Ser Ala Gly Val Gln Glu Ala
            20                  25                  30

Trp Ala Val Leu Gln Glu Ser Asn Val Pro Asp Leu Val Trp Thr Arg
        35                  40                  45

Cys Asn Gly Gly His Trp Ile Ala Thr Arg Gly Gln Leu Ile Arg Glu
    50                  55                  60

Ala Tyr Glu Asp Tyr Arg His Phe Ser Ser Glu Cys Pro Phe Ile Pro
65                  70                  75                  80

Arg Glu Ala Gly Glu Ala Tyr Asp Phe Ile Pro Thr Ser Met Asp Pro
                85                  90                  95

Pro Glu Gln Arg Gln Phe Arg Ala Leu Ala Asn Gln Val Val Gly Met
            100                 105                 110

Pro Val Val Asp Lys Leu Glu Asn Arg Ile Gln Glu Leu Ala Cys Ser
        115                 120                 125

Leu Ile Glu Ser Leu Arg Pro Gln Gly Gln Cys Asn Phe Thr Glu Asp
    130                 135                 140

Tyr Ala Glu Pro Phe Pro Ile Arg Ile Phe Met Leu Leu Ala Gly Leu
145                 150                 155                 160

Pro Glu Glu Asp Ile Pro His Leu Lys Tyr Leu Thr Asp Gln Met Thr
                165                 170                 175

Arg Pro Asp Gly Ser Met Thr Phe Ala Glu Ala Lys Glu Ala Leu Tyr
            180                 185                 190

Asp Tyr Leu Ile Pro Ile Ile Glu Gln Arg Arg Gln Lys Pro Gly Thr
        195                 200                 205

Asp Ala Ile Ser Ile Val Ala Asn Gly Gln Val Asn Gly Arg Pro Ile
    210                 215                 220

Thr Ser Asp Glu Ala Lys Arg Met Cys Gly Leu Leu Leu Val Gly Gly
225                 230                 235                 240

Leu Asp Thr Val Val Asn Phe Leu Ser Phe Ser Met Glu Phe Leu Ala
```

245 250 255

Lys Ser Pro Glu His Arg Gln Glu Leu Ile Glu Arg Pro Glu Arg Ile
260 265 270

Pro Ala Ala Cys Glu Glu Leu Leu Arg Arg Phe Ser Leu Val Ala Asp
275 280 285

Gly Arg Ile Leu Thr Ser Asp Tyr Glu Phe His Gly Val Gln Leu Lys
290 295 300

Lys Gly Asp Gln Ile Leu Leu Pro Gln Met Leu Ser Gly Leu Asp Glu
305 310 315 320

Arg Glu Asn Ala Cys Pro Met His Val Asp Phe Ser Arg Gln Lys Val
325 330 335

Ser His Thr Thr Phe Gly His Gly Ser His Leu Cys Leu Gly Gln His
340 345 350

Leu Ala Arg Arg Glu Ile Ile Val Thr Leu Lys Glu Trp Leu Thr Arg
355 360 365

Ile Pro Asp Phe Ser Ile Ala Pro Gly Ala Gln Ile Gln His Lys Ser
370 375 380

Gly Ile Val Ser Gly Val Gln Ala Leu Pro Leu Val Trp Asp Pro Ala
385 390 395 400

Thr Thr Lys Ala Val
405

<210> 6
<211> 1578
<212> DNA
<213> Pseudomonas putida

<400> 6
ctgcaggatc gttatccgct ggccgatctg atcacccagc gtttttccat cgacgaggcc 60
agcaaggcac ttgaactggt caaggcagga gcactgatca aacccgtgat cgactccact 120
ctttagccaa cccgcgttcc aggagaacaa caacaatgac gactgaaacc atacaaagca 180
acgccaatct tgcccctctg ccaccccatg tgccagagca cctggtattc gacttcgaca 240
tgtacaatcc gtcgaatctg tctgccggcg tgcaggaggc ctgggcagtt ctgcaagaat 300
caaacgtacc ggatctggtg tggactcgct gcaacggcgg acactggatc gccactcgcg 360
gccaactgat ccgtgaggcc tatgaagatt accgccactt ttccagcgag tgcccgttca 420
tccctcgtga agccggcgaa gcctacgact tcattcccac ctcgatggat ccgcccgagc 480
agcgccagtt tcgtgcgctg gccaaccaag tggttggcat gccggtggtg gataagctgg 540

agaaccggat ccaggagctg gcctgctcgc tgatcgagag cctgcgcccg caaggacagt 600
gcaacttcac cgaggactac gccgaaccct tcccgatacg catcttcatg ctgctcgcag 660
gtctaccgga agaagatatc ccgcacttga aatacctaac ggatcagatg acccgtccgg 720
atggcagcat gaccttcgca gaggccaagg aggcgctcta cgactatctg ataccgatca 780
tcgagcaacg caggcagaag ccgggaaccg acgctatcag catcgttgcc aacggccagg 840
tcaatgggcg accgatcacc agtgacgaag ccaagaggat gtgtggcctg ttactggtcg 900
gcggcctgga tacggtggtc aatttcctca gcttcagcat ggagttcctg gccaaaagcc 960
cggagcatcg ccaggagctg atcgagcgtc ccgagcgtat tccagccgct tgcgaggaac 1020
tactccggcg cttctcgctg gttgccgatg gccgcatcct cacctccgat tacgagtttc 1080
atggcgtgca actgaagaaa ggtgaccaga tcctgctacc gcagatgctg tctggcctgg 1140
atgagcgcga aaacgcctgc ccgatgcacg tcgacttcag tcgccaaaag gtttcacaca 1200
ccacctttgg ccacggcagc catctgtgcc ttggccagca cctggcccgc cgggaaatca 1260
tcgtcaccct caaggaatgg ctgaccagga ttcctgactt ctccattgcc ccgggtgccc 1320
agattcagca caagagcggc atcgtcagcg gcgtgcaggc actccctctg gtctgggatc 1380
cggcgactac caaagcggta taaacacatg ggagtgcgtg ctaagtgaac gcaaacgaca 1440
acgtggtcat cgtcggtacc ggactggctg gcgttgaggt cgccttcggc ctgcgcgcca 1500
gcggctggga aggcaatatc cggttggtgg gggatgcgac ggtaattccc catcacctac 1560
caccgctatc caaagctt 1578

<210> 7
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:synthetic

<400> 7
ccatggacgc tatcagcatc gttgccaac 29

<210> 8
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:synthetic

<400> 8
ccggcttctg cctgcgttgc tcga 24

<210> 9
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:synthetic

<400> 9
ccatggacaa ccctcaggac tttattgat 29

<210> 10
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:synthetic

<400> 10
ccattgattc ttggtgttct tttact 26

<210> 11
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:synthetic

<400> 11
gcatgaacaa ccctcaggac tttattga 28

<210> 12
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:synthetic

<400> 12
ccggcttctg cctgcgttgc tcg 23

<210> 13
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:synthetic

<400> 13
catcaccatc accatcactg aagaagagca gatggcctgg c 41

<210> 14
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:synthetic

<400> 14
gacaggaatg aagcacagct ggta 24

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/16979

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : B09B 3/00; C12N 1/00, 5/10, 9/02, 15/53; 15/63; C12P 1/00, 7/02

US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.2; 23.7; 435/41, 56, 57, 58, 59, 61,125, 189, 262.5, 69.1, 320.1, 252.3, 254.11, 325, 410

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,114,852 A (YABUSAKI et al.) 19 May 1992, entire document.	1-44
A	US 5,240,831 A, (H.J. BARNES) 31 August 1993, entire document.	1-44
A	O'KEEFE et al. Occurrence and biological function of cytochrome P450 monooxygenases in the actinomycetes. Molecular Microbiology. 1991. Vol. 5, No. 9, pages 2099-2105, entire document.	25-44
A	OKUDA et al. Recent progress in enzymology and molecular biology of enzymes involved in vitamin D metabolism. Journal of Lipid Research. 1995. Vol. 36, pages 1641-1652, entire document.	25-34

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 18 SEPTEMBER 1998	Date of mailing of the international search report 23 OCT 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer GABRIELE E. BUGAISKY Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/16979

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, P	SHIMOJI et al. Design of a Novel P450: A Functional Bacterial-Human Cytochrome P450 Chimera. Biochemistry. 1998. Vol. 37, No. 25, pages 8848-8852, entire document.	1-44

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/16979

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

536/23.2; 435/41, 189, 262.5, 69.1, 320.1, 252.3, 254.11, 325, 410

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, STN-CAS files Registry, Caplus, Biotechds, Derwent WPI; A-geneseq32, pir56, swissprot35, sptrembl16
search terms: cytochrome p450, fusi?, chimer?, bacter?, prokaryot?, eukaryot?, yeast mammalian, pseudomonas, putida, cyp2c9